 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS

Help




> Getting Started with the Data

Slide
\$cpInfoCurrentSlide
of \$cpInfoSlideCount

Getting Started with the Data

 [Download Transcript](#)  [Download Slides](#)  [Download Glossary](#)

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS


Help

> Getting Started with the Data > Module Objectives

00:00:38

Module Objectives

- Describe the resources available within the restricted-use CD:
 - Contents
 - File formats
 - File structure
 - Variable information and conventions
- Describe the resources available to learn more about the study, the data, and the data files

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS


Help

> Getting Started with the Data > Resources Available Within the Restricted-Use CD

00:01:43

Resources available within the restricted-use CD

- README file
 - Most important resource on the CD; always start with the README file
 - Information about CD contents and structure
 - Installation information
- Codebook
- Reference materials
 - Interview facsimiles, source documents, etc.
- Syntax used to derive variables
- Reports
 - Methodological and/or analytical

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS


Help

> Getting Started with the Data > Restricted-Use Micro-Level Data File Formats

00:01:03

Restricted-Use Micro-Level Data File Formats

- Depending on study, format of data may be:
 - Files in .CSV form accompanied by SAS, STATA, and SPSS syntax to create data files for each software package
 - Formatted files in SAS, STATA, and SPSS
 - Unformatted files accompanied by Electronic Codebook (ECB or ECBW) software to create SAS or SPSS syntax and files
- Consult the README file for installation details

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS


Help

> Getting Started with the Data > Micro-Level Data File Structure

00:01:19

Micro-Level Data File Structure

- CD contains multiple files that broadly fall into three categories:
 - Derived
 - Source
 - Weights
- All files contain ID variables to allow merging records across files

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS


Help

> Getting Started with the Data > Micro-Level Data File Structure: Derived Files

00:00:47

Micro-Level Data File Structure: Derived Files

- Ready to use files for conducting analyses
 - Files that PowerStats use
- Built from source files to:
 - Pick the most reliable source for a given item
 - Fill in missing information from the interview
 - Generate summary indicators (ex: ratio of debt to earnings)
- Include edits and imputations
- Missing values labeled
- Codebooks and syntax on the CD outline how each variable was derived

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS


Help

> Getting Started with the Data > Micro-Level Data File Structure: Source Files

00:00:60

Micro-Level Data File Structure: Source Files

- Records received from various data collection efforts
 - Student interviews (multiple for longitudinal studies)
 - Institutional information from IPEDS
 - Student information provided by the institution, also known as Computer Aided Data Extraction, or CADE
 - Administrative data sources, such as:
 - Central Processing System (CPS) information on financial aid application (FAFSA)
 - National Student Loan Data System (NLSDS) information on student loans
 - Pell grant recipient
 - College entrance exam scores
 - National Student Clearinghouse information on attendance
 - College transcripts
- It is highly unlikely that you will need to consult a source file for analysis.

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS


Help

> Getting Started with the Data > Micro-Level Data File Structure: Weight Variables/Files

00:00:50

Micro-Level Data File Structure: Weight Variables/Files

- [Replicate weights](#) may be in derived files or in separate weights files on the restricted-use CD
 - Weight files contain
 - Replicate weights
 - Weight components
 - Derived files typically include, at a minimum:
 - Final analysis weight
 - Design indicators (strata and primary sampling unit)

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS

Help


> Getting Started with the Data > Micro-Level Data File Structure: File Information

00:00:58

Micro-Level Data File Structure: File Information

- For each file, the README file provides:
 - Content summary
 - File name
 - Number of cases
 - Number of records*
- File name usually includes whole word or abbreviated form of “DERIVED”

* May not match number of cases if multiple records exist per case

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS


Help

> Getting Started with the Data > Micro-Level Data File Structure: File Information

00:00:24

Micro-Level Data File Structure: File Information

- May be more than one derived file for each study:
 - Longitudinal studies: each follow-up and panel as a whole
 - Student-level school derived file
 - Student-level transcript derived file

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS

Help


> Getting Started with the Data > Micro-Level Data Variable Information

00:01:13

Micro-Level Data Variable Information

- Codebook outlines following details for variables:
 - Name*
 - Description
 - Label
 - Subject area
 - Detail about who the variable applies to
 - Unweighted frequency distribution including [reserve codes](#) for missing valid values, such as:
 - -3: Item was skipped for individual
 - -9: Item was missing for individual

* ECBW users should use caution: the same variable name may appear in multiple files. Be sure you are always selecting your variables from the final derived file.

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS


Help

> Getting Started with the Data > Resources Available from Other Sources

00:01:42

Resources Available from Other Sources

- To learn more about the study, the data, and the data file
 - Visit the main study pages for study overviews, publications, and staff contact information
 - [National Postsecondary Student Aid Study \(NPSAS\)](#)
 - [Beginning Postsecondary Students Longitudinal Study \(BPS\)](#)
 - [Baccalaureate and Beyond Longitudinal Study \(B&B\)](#)
 - Visit the [DataLab](#) to search published tables and create custom tables and analyses in QuickStats and PowerStats


 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS

Help

> Getting Started with the Data > Summary

00:00:23

 INSTITUTE OF
EDUCATION SCIENCES

NATIONAL CENTER FOR
EDUCATION STATISTICS

Help

> Getting Started with the Data > Resources

00:00:18

Module Resources

- [Building derived variables example](#)
- [Replicate weights](#)
- [Reserve codes](#)
- [NPSAS](#), [BPS](#), and [B&B](#) homepages
- [DataLab](#)